DOCUMENT RESUME

ED 476 506 SO 034 860

AUTHOR	McEwan, Patrick J.; Jimenez, Wilson
TITLE	Indigenous Students in Bolivian Primary Schools: Patterns and Determinants of Inequities. A World Bank Study.
INSTITUTION	World Bank, Washington, DC.
PUB DATE	2002-08-27
NOTE	46p.
AVAILABLE FROM	The World Bank, 1818 H Street, N.W., Washington, DC 20433. Tel: 800-645-7247 (Toll Free); Fax: 703-661-1501; Web site: http://www.worldbank.org/ .
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	Cultural Context; Developing Nations; Educational Research; Elementary Education; *Equal Education; Foreign Countries; Grade 3; Grade 6; *Indigenous Populations; Outcomes of Education
IDENTIFIERS	*Bolivia; Empirical Research

ABSTRACT

This paper examines the patterns and determinants of inequities between indigenous and non-indigenous students in Bolivian primary (elementary) schools, using a 1997 survey of third and sixth graders. Analysis shows that, on average, indigenous students are of lower socioeconomic status, and their schools appear to have fewer instructional materials and poorer infrastructure. There is notable student segregation by linguistic status. Indigenous students score 0.28-0.47 standard deviations lower than non-indigenous student on tests of mathematics and Spanish achievement. Approximately two-thirds of the difference between indigenous student and non-indigenous student scores is explained by the unequal distribution of family and school variables. The remaining difference in scores may be explained by the unique linguistic background of students, or by unobserved features of families and schools. In the early grades, Bolivia scored below the regional mean of 11 Latin American countries. Bringing indigenous students to parity with others is the first part of a longer journey. (Contains 24 notes, 14 tables, 2 figures, and 41 references.) (Author/BT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 476 506

A WORLD BANK STUDY

Indigenous Students in Bolivian Primary Schools
Patterns and Determinants of Inequities

Patrick J. McEwan and Wilson Jiménez

August 27, 2002

SO 034 860

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

J. Feather

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Table of Contents

Foreword vi
Acknowledgements viii
Abstract x
1. Introduction 1
2. The Context of Bolivian Schooling 3
2.1 The Extent of Linguistic Diversity *3*
2.2 Linguistic Diversity and Formal Schooling *5*
3. Prior Research 8
4. Empirical Approach 10
4.1 The Distribution of Family and School Characteristics *10*
4.2 Patterns and Determinants of Achievement Differences *11*
5. Data 14
6. The Distribution of Family and School Characteristics 16
6.1 Family Characteristics *16*
6.2 Segregation by Indigenous Status *21*
6.3 School and Teacher Characteristics *22*
7. Patterns and Determinants of Achievement Differences 25
7.2 The Determinants of Academic Achievement *32*
7.3 Comparing Quechua and Aymara Students *33*
7.4 Interpretations *36*
8. Summary and Conclusions 38
References 40

List of Tables

Table 1 Evolution of Language Skills in the Population Aged 6 and Over 3
Table 2 Language Skills in Different Age Groups, 1997 5
Table 3 Variable Definitions 14
Table 4 Description Statistics, Students and Family Variables 16
Table 5 Description of Student and Family Variables among Groups, Probit Regression 18
Table 6 Description Statistics, Student and Family Variables 22
Table 7 Distribution of School and Teacher Variables Across Linguistic Groups, Probit Regressions 24
Table 8 Determinants of Third-Grade Spanish Achievement, 1997 25
Table 9 Determinants of Sixth-Grade Spanish Achievements. 1997 27
Table 10 Determinants of Third-Grade Mathematics Achievement, 1997 28
Table 11 Determinant of Sixth-Grade Mathematics Achievement, 1997 29
Table 12 Summary of Achievement Gaps Between Indigenous and Non-Indigenous Students, Expressed in Standard Deviations 30
Table 13 Alternative Specifications of the Determinants of Spanish Achievement, 1997 32

Table 14 Alternative Specifications of the Determinants of Mathematics Achievement, 1997 33

List of Figures

Figure 1 Distribution of Indigenous Students Across Primary Schools 19
Figure 2 Kernel Densities of Indigenous and Non-Indigenous Achievement 23

Foreword

Acknowledgements

This study was financed by the Partnership on Sustainable Strategies for Girls' Education, an international , interagency group dedicated improving educational opportunities for girls in the developing world. Partner agencies include the World Bank, the Department for International Development (U.K.), UNICEF, and the Rockerfeller Foundation.

The authors of this paper are Patrick J. McEwan[1] of the University of Illinois at Urbana-Champaign, and Wilson Jimenez[2] of Unidad de Análisis de Políticas Económicas. The views and conclusions expressed in this document are those of the authors and should not be attributed to the Partnership or its member agencies.

The authors gratefully acknowledge input and reviews provided by Xavier Albó, María Luisa Talavera, Miguel Urquiola, Miguel Vera, Donald Winkler and Terry Allsop, during the preparation of this paper. Formatting of the paper for production was undertaken by Ian Conachy and Benjamin Crow.

[1] Department of Educational Policy Studies; 360 Education Building; 1310 S. Sixth St.; Champaign IL 61820; USA; email: mcewan@uniuc.edu

[2] Ministerio de Hacienda; Palacio de Comunicaciones, Piso 18; La Paz; Bolivia; email: wjimenez@udape.gov.bo

Abstract

This paper examines the patterns and determinants of inequities between indigenous and non-indigenous students in Bolivian primary schools, using a 1997 survey of third and sixth graders. The analysis shows that, on average, indigenous students are of lower socioeconomic status, and their schools appear to have fewer instructional materials and poorer infrastructure. There is notable student segregation by linguistic status. Indigenous students score 0.28-0.47 standard deviations lower than non-indigenous students on tests of mathematics and Spanish achievement. Approximately two-thirds of the difference between indigenous students and non-indigenous students scores is explained by the unequal distribution of family and school variables. The remaining difference in scores may be explained by the unique linguistic background of students, or by unobserved features of families and schools.

1. Introduction

Spanish is spoken by roughly 6 million of the 6.7 million Bolivians over the age of 5, but at least half the population speaks a language other than Spanish. Conservatively, 1.8 million speak Quechua on a regular basis, 1.2 million speak Aymara, and 33,000 speak Guaraní.[3] A range of other indigenous languages—at least 30—are spoken by a far smaller number of individuals.

There are compelling—but poorly documented—reasons to expect that children of indigenous parents are at greater risk of educational failure than other children.[4] First, indigenous adults have, on average, less schooling and lower incomes than other adults (Psacharopoulos 1993; Chiswick, Patrinos, and Hurst 2000; Kelley 1988; Wood and Patrinos 1994). Both variables are common indicators of the quality of the educational environment in the home. Second, indigenous families are more likely to live in rural areas or in marginalized neighborhoods of urban areas, where public schools may have fewer and lower quality instructional resources. Third, Bolivian public schooling has traditionally ignored the diverse linguistic backgrounds of its students. The norm was to impose Spanish usage at all levels of schooling—regardless of students' or even teachers' knowledge of the language—and sometimes even punish the use of indigenous languages (Plaza and Albó 1989; Comitas 1972). To be fair, government-led education reforms—promulgated in 1994 and implemented gradually since then—have promoted a range of alternative approaches, including bilingual education (Hornberger and King 1996; Hornberger 2000). As yet, these have been subjected to little rigorous evaluation.

The prior arguments have led most to assume—bolstered by a mountain of case studies and anecdotal observation—that indigenous students are disadvantaged, relative to other students. Yet, there is a lacuna in the empirical work of academics and policy analysts. Almost nothing of a systematic nature is known about the inequities that exist in the schooling outcomes of indigenous and non-indigenous students and of girls and boys of indigenous and non-indigenous backgrounds. Further, little is known about the determinants of such inequities and consequently about their remedies.

This study attempts to provide some redress. It relies on a survey of academic achievement among third and sixth graders in 1997, administered by the *Sistema de Medición de la Calidad de la Educación* (hereafter SIMECAL). The survey contains a variety of background information on students, parents, teachers, and principals. Using the SIMECAL data, this paper pursues three objectives.

---

[3] For general descriptions of linguistic diversity in Bolivia, see Grinevald (1998) and Albó (1995a, 1998). The cited figures are authors' estimates from the 1997 *Encuesta Nacional de Empleo*. For further description of these data and caveats associated with their use, see Section 2.

[4] This should not be interpreted to mean that non-indigenous children have high achievement. In fact, evidence suggests that the overall language and mathematics achievement of Bolivian children is below the regional mean of 11 Latin American countries (Willms and Somers 2001).

First, it describes the patterns of inequities in academic achievement between indigenous and non-indigenous students, focusing on students in families where Quechua and Aymara are spoken. The analysis also describes inequities between genders, and between urban and rural students.

Second, it describes patterns of inequities in the distribution of family characteristics such as parental schooling, and basic school characteristics such as instructional materials and teacher qualifications.

Third, it assesses whether inequities in achievement are attributable to the inequitable distribution of family and school characteristics, or whether the gap is not explained by these factors. The latter case might imply a stronger explanatory role for students' unique linguistic backgrounds.

Three caveats should be noted at the outset. First, this study focuses exclusively on girls and boys children who are observed to attend school since this study does not consider the outcomes of children outside the formal school system. Other evidence indicates that indigenous children in Bolivia are more likely to enter school late, to repeat grades, and to drop out of school early (Patrinos and Psacharopoulos 1996; Jimenez and McEwan 2001) it is possible that this study yields conservative estimates of the achievement gap between the entire population of indigenous and non-indigenous students.

Second, this study focuses exclusively on two measures of schooling outcomes: academic achievement in Spanish and mathematics. There are clearly other important outcomes; not least in this context is the literacy and achievement of students in languages other than Spanish.

Third, this study defines a student as "indigenous" if he or she uses an indigenous language to communicate in the home (and our repeated use of the term should be interpreted with this in mind).

The paper is organized as follows: Section 2 describes the context of schooling in Bolivia, including the extent of linguistic diversity among students and provides a brief history of education reforms that address this diversity; Section 3 reviews the extant research on the schooling outcomes of indigenous and non-indigenous individuals in Latin America; Sections 4 and 5 review the empirical approach and data, respectively, that are used to explore the main research questions; Section 6 describes the distribution of family and school characteristics across groups; and Section 7 describes patterns and determinants of inequities in achievement. The final section summarizes the findings and draws conclusions.

2. The Context of Bolivian Schooling

A defining characteristic of Bolivian schooling is the linguistic diversity that it must confront. Hence, this section describes the extent of that diversity, using recent data. It further describes the educational reforms and policies that have attempted to address this diversity.

2.1 The Extent of Linguistic Diversity

The three principal indigenous languages in Bolivia are Quechua, Aymara, and Guaraní.[5] Quechua and Aymara are by far the most widely spoken; Guaraní and other indigenous languages have considerably fewer speakers.[6] This paper focuses on the Quechua and Aymara populations in Bolivia, for a very practical reason. All of the data used in this paper are samples drawn from the Bolivian population. Because smaller indigenous groups were not over-sampled, it is difficult to make general statements about these groups with any degree of precision. Their exclusion here should not be interpreted as a statement about their relative importance.

Table 1 provides data on the relative size of the Aymara and Quechua populations, considering ages 6 and over, both in absolute numbers and as a share of the total population. The 1976 and 1992 data are drawn from the census, where language background is assessed with a simple question: "Do you speak Quechua (or Spanish or Aymara)?" Respondents are allowed to indicate that they speak more than one language. The 1997 data are drawn from a nationally representative household survey—the *Encuesta Nacional de Empleo*—in which individuals report the languages that they speak "regularly."

[5] Various dialects of Quechua, sometimes referred to by other names, are widely spoken in Ecuador, Peru, and, to a lesser extent, Argentina. Aymara is also spoken in southern Peru and northern Chile. Guaraní is spoken by the vast majority of Paraguayans and in parts of Argentina and Brazil. For additional details, see Grinevald (1998) and the citations therein.

[6] For sociolinguistic analyses of the Bolivian Guaraní population, see López (1997) and Albó (1995b).

Table 1. Evolution of Language Skills in the Population Aged 6 and Over

	Number ('000)			*Percent*		
	1976	*1992*	*1997*	*1976*	*1992*	*1997*
Knows Spanish	3,210	4,594	5,963	78.8	87.4	89.5
			(317)			(1.0)
Knows Quechua	1,594	1,806	1,780	39.7	34.3	26.7
			(139)			(1.4)
Knows Aymara	1,156	1,238	1,228	28.8	23.0	18.4
			(75)			(1.5)
Knows only Spanish	1,508	2,203	3,581	36.3	41.7	53.7
			(245)			(1.7)
Knows only Quechua	569	428	440	13.7	8.1	6.6
			(61)			(0.8)
Knows only Aymara	315	169	166	7.6	3.2	2.5
			(24)			(0.4)

Note: 1997 observations are weighted to account for unequal probabilities of selection into the sample. Standard errors, adjusted for the clustered sample design, are in paraentheses.

Sources: 1976 and 1992 estimates are from the census (Albó 1995a). 1997 estimates are from the *Encuesta Nacional de Empleo* and authors' calculations.

There are well-known challenges to measuring language skills in this manner.[7] In Bolivia, a particular dilemma is that Spanish is perceived as a high-prestige language, while indigenous languages are typically accorded less prestige.[8] Thus, respondents may overstate their knowledge or usage of Spanish and understate their knowledge of indigenous languages. In general, therefore, the following estimates of indigenous language use should be viewed as a lower bound.

With this caveat in mind, Table 1 supports four broad conclusions. First, the great majority of Bolivians have at least some knowledge of Spanish, and this rate steadily increased between 1976 and 1997. Currently, about 90 percent of Bolivians use Spanish regularly. Second, rates of Spanish monolingualism have marched steadily upward, reaching 54 percent in 1997. Third, a declining percentage of the population speaks Quechua or Aymara. However, the percentages are still quite substantial, with 27 percent using Quechua regularly and 18 percent using Aymara. Fourth, the rates of monolingualism in an indigenous language declined precipitously during the period in question, to the point where less than 10 percent of the population speaks only Aymara or Quechua.

However, the 1997 figures may be slightly misleading because of the wording of the question. It is possible that some individuals do not use a language "regularly," but still "know" how to speak that language. Therefore, the 1997 data might overstate the degree of language shift. For this added reason, one should regard the 1997 estimates of indigenous language use as lower bounds.

Using the 1997 data, Table 2 reports the same analyses, disaggregated by age cohorts and gender. Although the data are cross-sectional, they are also suggestive that indigenous language use has declined over time. Older cohorts are considerably more likely to speak Aymara or Quechua. Among

[7] For a review of issues, see Romaine (2000).

[8] See Albó (1995a). Empirical studies of language attitudes in Bolivia and Peru support this conclusion (Wölck 1973; Briggs 1979).

BEST COPY AVAILABLE

the two oldest cohorts, for example, between 20 and 30 percent are monolingual in Aymara or Quechua. Among the two youngest cohorts, monolingualism in Aymara appears to be almost nil, while less than 5 percent are monolingual in Quechua.

These patterns are generally similar across genders, although language shifts appears to have occurred even more rapidly for females. Among older cohorts, females use Spanish less than males and are more likely to use only an indigenous language. Among younger cohorts, gender inequalities in language usage have all but disappeared.

Table 2. Language Skills in Different Age Groups, 1997

	5-10	*11-15*	*16-20*	*21-30*	*31-40*	*41-50*	*51-60*	*61-70*
Panel A: Entire Sample (%)								
Knows Spanish	93.1	95.8	95.4	92.3	89.5	85.2	78.8	67.3
Knows Quechua	18.8	21.5	22.1	25.1	30.0	33.2	34.5	43.3
Knows Aymara	11.6	13.7	12.6	18.3	23.1	23.7	26.4	28.1
Knows only Spanish	67.4	63.1	64.0	55.1	46.1	42.2	38.7	30.2
Knows only Quechua	4.4	2.9	3.3	4.8	6.9	9.9	12.1	19.4
Knows only Aymara	1.0	0.2	0.3	1.3	2.2	3.4	7.5	10.3
Panel B: Females (%)								
Knows Spanish	92.6	94.6	94.7	91.6	87.5	81.2	71.8	63.3
Knows Quechua	18.8	22.1	22.4	24.0	29.7	33.7	35.4	43.9
Knows Aymara	11.6	13.4	14.1	18.3	23.5	22.6	27.1	28.3
Knows only Spanish	67.6	62.6	62.1	56.1	46.1	42.7	37.2	29.7
Knows only Quechua	4.4	3.9	3.7	5.1	8.5	12.6	15.0	21.4
Knows only Aymara	1.2	0.4	0.5	1.4	2.5	4.3	11.4	12.6
Panel C: Males (%)								
Knows Spanish	93.4	97.0	96.2	9\.1	91.6	89.5	85.7	71.7
Knows Quechua	18.8	20.9	21.8	2\.2	30.3	32.6	33.6	42.6
Knows Aymara	11.6	13.9	11.0	18.2	22.7	24.8	25.7	29.2
Knows only Spanish	67.3	63.5	65.9	53.9	46.2	41.6	40.1	30.8
Knows only Quechua	4.4	2.1	3.0	4.5	5.1	7.0	9.2	17.2
Knows only Aymara	0.8	0.1	0.1	1.2	1.9	2.4	3.8	7.8
Panel D: Difference (C-B)(%)								
Knows Spanish	0.8	2.4	1.5	1.5	4.1	8.3	13.9	8.4
Knows Quechua	0.0	-1.2	-0.6	2.2	0.6	-1.1	-1.8	-1.3
Knows Aymara	0.0	0.5	-3.1	-0.1	-0.8	2.2	-1.4	0.9
Knows only Spanish	-0.3	0.9	3.8	-2.2	0.1	-1.1	2.9	1.1
Knows only Quechua	0.0	-1.8	-0.7	-0.6	-3.4	-5.6	-5.8	-4.2
Knows only Aymara	-0.4	-0.3	-0.4	-0.2	-0.6	-1.9	-7.6	-4.8

Note: 1997 observations are weighted to account for unequal probabilities of selection into the sample.
Source: 1997 *Encuesta Nacional de Empleo* and authors' calculations.

2.2 Linguistic Diversity and Formal Schooling

Historically, Bolivia's system of formal schooling has not considered issues of linguistic diversity in the classroom. Between 1976 and 1990, however, four programs implemented some form of

BEST COPY AVAILABLE

bilingual education.[9] In general, these were not implemented on a large scale, and they did not prove to be sustainable.

First, the *Proyecto Educativo Rural* was initiated in 1976 by the Ministry of Education with support from USAID. It targeted 22 schools serving the Quechua-speaking population of Cochabamba. Second, the *Proyecto Educativo Integrado del Altiplano* was started by the Ministry of Education in 1978 with a loan from the World Bank. It was implemented in 14 schools located in rural and predominantly Aymara-speaking regions of La Paz.

A third project was begun in 1981 by the Catholic Church with some financial support from the Ministry of Education. The *Proyecto Texto Rural Bilingüe* was initially directed at Aymara populations and later at Quechua speakers. Fourth, UNICEF and the Ministry of Education operated the *Proyecto de Educación Intercultural y Bilingüe* between 1990 and 1994. It functioned in a larger number of schools (more than 100), and variants were designed for a range of indigenous groups including speakers of Aymara, Quechua, and Guaraní.[10]

Notwithstanding these efforts, the official policies of the Ministry of Education did not seriously contemplate bilingual education until 1994. In that year, an Education Reform Law stipulated that Bolivian schools—including but not limited to those serving indigenous students—would undergo a sweeping institutional and curricular transformation. Among other goals, the reform intended to deliver an entirely new curriculum, supported by newly designed instructional materials, teacher training, and community participation. A key element of the curriculum is its emphasis on bilingual and intercultural education (Hornberger and King 1996). Students in predominantly indigenous schools are eligible to receive instruction in both Spanish and their indigenous language, while students in monolingual Spanish schools are supposedly eligible to take an indigenous language as an additional subject.

The reform's implementation proceeded gradually (and, indeed, is ongoing). In 1996, the reform was applied in an initial group of schools in the first grade. In subsequent years, the reform was applied to additional schools and to successively higher grades in already participating schools. Regarding the bilingual component, activities proceed on several fronts in the Ministry of Education. There is ongoing development of instructional materials in several languages. In addition, pedagogical advisors *("asesores pedagógicos")* have been trained at several teacher institutes that are now specifically charged with supporting the bilingual, intercultural reforms.

The authors are not aware of attempts to rigorously evaluate the impact of these programs and reforms on student outcomes such as academic achievement in Spanish, mathematics, or an indigenous language. That is not to say that some authors have not measured and reported these outcomes for groups of students that attended participating schools. Some assessments of a UNICEF project have been undertaken (Muñoz, 1997 and SIMECAL, 2000). However, while these studies describe the competencies of students, they cannot tell us whether these competencies would have been different (or

[9] For a review of the three earliest programs, see Plaza and Albó (1989) and Briggs (1982, 1985).

[10] Muñoz (1997) reports the findings of an external evaluation, contracted by UNICEF. López (1997) argues that the program may have been particularly effective among Guaraní populations due to the extensive participation of local communities.

the same) in the absence of the program. Ultimately, this is the causal, evaluative question that is most interesting: Did the program successfully increase student outcomes?[11]

It is important to note that the present study will not attempt to provide a conclusive answer to that question. The reason for this is largely practical. The recent Bolivian reforms were not implemented until 1996, and then only in the beginning grades of a limited number of schools. The data used in this paper, described in Section 4, were collected in 1997 in grades 3 and 6—effectively preceding the reform.

[11] To answer this question, one must compare participating students with a roughly similar group of nonparticipating students. This similarity is best obtained by randomly assigning students to each group. If a randomized, experimental evaluation cannot be implemented—which clearly holds in the present circumstances—there are a range of quasi-experimental and nonexperimental methods that can be used to identify the causal impact of the program on achievement (for a description, see Angrist and Krueger (1999)). To date, none of these methods have been applied to the impact evaluation of bilingual education in Bolivia.

BEST COPY AVAILABLE

3. Prior Research

This section briefly reviews empirical research that compares indigenous and non-indigenous schooling outcomes such as test scores. In Latin America, only three studies were located.[12] Rivarola, Corvalán, and Zuniga (1977) analyze fourth-grade reading scores in Paraguay. They found that Spanish monolinguals scored higher than those of both bilingual students and those who are monolingual in Guaraní. However, the relationship holds only for urban areas.

More recently, an analysis of fourth-grade achievement in Peru suggests that Quechua speakers score worse than Spanish speakers, even after controlling for a range of family and student variables; there is no statistically significant difference between Aymara and Spanish speakers (World Bank 2001). However, the results are difficult to interpret for two reasons. First, the standard deviation of the dependent variable is not reported. Hence, it is difficult to know whether the magnitudes of the observed differences in test scores are large or small. Second, Aymara speakers are only 3 percent of the sample, suggesting that statistical power may be insufficient to detect a negative effect for this group. In fact, the negative regression coefficients for Quechua and Aymara speakers are roughly equal.[13]

Perhaps the best study is an unpublished thesis by Vera (1998), who analyzes a 1992 sample of fourth graders from the Bolivian cities of La Paz and El Alto.[14] On average, bilingual students score 7.1 points less than Spanish monolinguals on a mathematics test; this represents a rather substantial effect size of 0.7 standard deviations. After controlling for a limited set of socioeconomic variables, the difference declines to 2.8 points, or 0.27 when expressed in standard deviations.[15] Note that indigenous students in these cities are largely Aymara (Albó 1995a), and the results should be interpreted accordingly.

The prior discussion was limited to short-term schooling outcomes such as academic achievement. However, a growing body of research compares long-term schooling outcomes such as labor market earnings (for a review, see Patrinos (2000)). In a study of Bolivia, using urban data from 1989, Wood and Patrinos (1994) find that indigenous earnings are 61 percent of nonindigenous earnings. However, 73 percent of this gap can be directly explained by differences in each group's observed characteristics such as the quantity of schooling received. The remaining gap may be due to discrimination in the labor market, but it is also potentially caused by unobserved productivity differences between indigenous and nonindigenous workers. These, in turn, may stem from differences in the quality of the education that was received. The empirical analysis of this paper will suggest that

[12] In contract, there a voluminous international literature that analyzes the academic outcomes of various ethnic and linguistic minorities. For example, the edited volume of Jencks and Phillips (1998) contains a wide array of articles on the magnitude and determinants of the black-white test score gap in the United States. ON the differences between Maori and white academic outcomes in New Zealand, see Fergusson, Lloyd, and Horwood (1991). Lavy (1998) compares the schooling outcomes of Arabs and Jews in Israel.

[13] No standard errors are reported for the coefficients. Thus, one cannot directly assess whether the coefficient are statistically different.

[14] In fact, these data were a pi' . for a larger data collection in 1997. The latter are the data used in the present analysis.

[15] These results are taken fro Appendix 2, Model 1 (Vera 1998).

BEST COPY AVAILABLE

unobserved schooling quality, proxied by test scores, could play some role in explaining the remaining earnings gap.

4. Empirical Approach

This section describes the empirical approach that will be used to answer three questions: (i) How are family and school characteristics distributed across indigenous and non-indigenous students and other groups, such as boys and girls; (ii) How is academic achievement distributed across the same groups? and (iii) What are the determinants of inequalities in achievement?

4.1 The Distribution of Family and School Characteristics

Assume that one is comparing the years of parental schooling of indigenous and nonindigenous students. The simplest approach is to estimate the mean schooling of each group, employing a t-test to assess whether the group means are statistically different. The same task could be accomplished with a regression model:

$$S_i = \beta_0 + \beta_1 I_i + \varepsilon_i \quad (1)$$

where S_i is the parental schooling of student i, I_i is a dummy variable measuring the indigenous status of each student (equal to 1 if the student is indigenous and 0 otherwise), and e_i is an error term. The constant β_0 provides an estimate of the mean parental schooling of nonindigenous students. β_1 provides an estimate of whether parental schooling of indigenous students is greater or smaller. Its standard error allows one to assess whether the difference is statistically different from zero.

In addition to indigenous status, one might explore how family and school characteristics vary in relation to gender and rural/urban location. Consider an expanded model of the following form:

$$S_i = \beta_0 + \beta_1 I_i + \beta_2 F_i + \beta_3 R_i + \varepsilon_i \quad (2)$$

where F_i is a dummy variable, equal to 1 for girls and 0 for boys. Likewise, R_i is a dummy variable, equal to 1 for students in rural areas and 0 for students in urban areas. To some extent, the inclusion of these variables, particularly R_i, may explain some of the disparities between indigenous and non-indigenous students, because the former are disproportionately located in rural areas.

One might refine the analysis even further by including interaction terms between indigenous status and the other variables:

$$S_i = \beta_0 + \beta_1 I_i + \beta_2 F_i + \beta_3 R_i + \beta_4 (I_i * F_i) + \beta_5 (I_i * R_i) + \varepsilon_i \quad (3)$$

Estimates of β_4 and β_5 allow for a more careful assessment of inequalities in family or school characteristics. For example, are gender gaps more or less pronounced among indigenous children? Similarly, is the gap between indigenous and non-indigenous students more pronounced in rural areas?

This paper's empirical analysis reports estimates of equation (3) for each of the family and school variables in the Bolivian data that will be described in Section 5. However, most of these variables are dichotomous, rather than continuous. Thus, the estimates will use maximum likelihood probit regressions to obtain coefficient estimates, rather than ordinary least squares. Moreover, instead of reporting coefficients from probit regressions, each analysis will report marginal effects that can be directly interpreted as the change in probability associated with a change in each independent variable.

4.2 Patterns and Determinants of Achievement Differences

The analysis of achievement employs a similar approach. Three alternate specifications are estimated for each dependent variable:

$$A_{ijk} = \beta_0 \beta_1 I_{ijk} + \varepsilon_{ijk} \tag{4}$$

$$A_{ijk} = \beta_0 + \beta_1 I_i + \beta_2 F_i + \beta_3 R_{ijk} + \varepsilon_i \tag{5}$$

$$A_{ijk} = \beta_0 + \beta_1 I_{ijk} + \beta_2 F_{ijk} + \beta_3 R_{ijk} + \beta_4 \left(I_{ijk} * F_{ijk}\right) + \beta_5 \left(I_{ijk} * R_{ijk}\right) + \varepsilon_{ijk} \tag{6}$$

where A_{ijk} is the academic achievement of student i in classroom j in school k (and other variables are the same as before). Coefficient estimates from these models provide a compact description of the differences in achievement between indigenous and nonindigenous students, between boys and girls, and between rural and urban students (as well as interactions).

To assess whether these differences can be explained by varying endowments of family and school characteristics, three additional models are estimated:

$$A_{ijk} = \ldots + \mathbf{X}_{ijk}\beta_6 + \beta_{ijk} \tag{7}$$

$$A_{ijk} = \ldots + \mathbf{X}_{ijk}\beta_6 + \mu_k + \varepsilon_{ijk} \tag{8}$$

$$A_{ijk} = \ldots \mathbf{X}_{ijk}\beta_6 + \mu_j + \varepsilon_{ijk} \tag{9}$$

where each ellipsis indicates the same constant and variables included in equation (6). Equation (7) adds a vector of family and student control variables ($\mathbf{X}_{ijk}$), including the socioeconomic status of families. The inclusion of these variables should modify the estimates of β_1-β_5, if family and student variables are correlated with achievement as well as indigenous status, gender, and rural/urban location. To some extent, one can interpret this to mean that family and student variables "explain" the achievement gaps between various groups. However, this interpretation should be made with caution because it is quite possible that family variables are capturing the influence of unmeasured school and classroom variables.

To assess whether group differences in achievement are further explained by the unequal distribution of school variables, equation (8) adds school fixed effects (μ_k)—essentially a dummy variable for each school. There are strengths and weaknesses to this approach. The primary strength is that it allows one to control for all variables that are constant within individual schools (such as infrastructure, principal characteristics, or other resource endowments), *whether or not they are*

observed by the researcher. Hence, the resulting estimates of achievement gaps between groups are entirely purged of the influence of school variables. The primary weakness is that one cannot estimate separate coefficients for individual school variables. As it turns out, this is no great limitation, because a substantial number of cases in the SIMECAL data are missing observations for school variables. Thus, the estimation of school coefficients would have implied a drastically reduced and perhaps unrepresentative sample.

Equation (8), the final model, includes classroom fixed effects (μ_j) instead of school fixed effects. Besides subsuming school variables, the inclusion of classroom fixed effects controls for variables that are constant across each classroom within a school (such as teacher characteristics, instructional materials in each classroom, and so on). Thus, the estimates of β_1-β_5 no longer reflect the influence of classroom variables.

The remaining differences in achievement between groups have several potential interpretations that are discussed in the results.

5. Data

To implement the empirical strategy, the paper employs data collected by the *Sistema de Medición de la Calidad de la Educación*. SIMECAL, a unit of the Ministry of Education, administered a national survey of academic achievement for the first time in 1997.[16] A sample of primary schools was drawn, and students in each school's third- and sixth-grade classrooms took achievement tests in Spanish and mathematics.[17] In addition, background surveys were filled out by students, parents, teachers, principals, and community leaders.

A limited set of these variables, reported in Table 3, are used in this paper's analysis. The key independent variable is INDIGENOUS, equal to 1 if a parent reports that an indigenous language is used to communicate in the home and zero otherwise. The SIMECAL data also contain a student-reported variable: "What languages do you speak?" This paper uses the parent-reported variable for two reasons. First, there are fewer missing cases. Second, one suspects that parents, especially compared with third graders, will provide answers that are less error prone. Notwithstanding this rationale, a sensitivity analysis suggested that estimates of achievement differences are not substantially different, depending on the variable used.

This general variable is replaced in some of the subsequent analyses by two more detailed variables: QUECHUA and AYMARA, indicating that the respective languages are spoken in the home.[18] It is worth noting that QUECHUA and AYMARA are not always mutually exclusive. That is, some families report that both indigenous languages are used in the home, although these only constitute about 1 percent of the sample.

The other family and student variables measure a range of socioeconomic characteristics, including parental education and availability of basic services such as electricity and sewer. Including the latter variables does not necessarily imply that these services directly influence achievement. However, they may serve as useful proxies for other variables, including family income and wealth. All of the family and student variables are dichotomous, rather than continuous variables.

Table 3 also describes a range of school and teacher variables. Although these variables are not used in the analysis of academic achievement—which employs school and classroom fixed effects—the next section will use these variables to assess how school and teacher characteristics are distributed across indigenous and non-indigenous students. They reflect simple indicators of school and teacher

[16] For other analyses of the 1997 data, see Urquiola (2000) and Urquiola, Jiménez, Talavera, and Hernany (2000), who use the third-grade sample, and Mizala, Romaguera, and Reinaga (1999), who use the sixth-grade sample.

[17] The sample was clustered in the sense that schools were the primary sampling unit, followed by students within schools. It is not clear whether the sample is self weighting or whether sample weights are required to obtain sample means that are appropriate estimates of population means. In any case, sample weights are not provided with the official data. Because of this, all sample means reported in this paper are unweighted and should only be used to draw inferences about population values with some caution.

[18] There are a very small number of parents—totaling about 1 percent of the sample—who report that Guarani or another is spoken in the home. These cases are excluded from the analysis.

BEST COPY AVAILABLE

quality, including the availability of instructional materials, teacher education, and the condition of school infrastructure. With one exception, the variables are dichotomous.

Table 3. Variable Definitions

Variable	*Definition*
Dependent variables:	
SPANISH	Spanish achievement test score
MATH	Mathematics achievement test score
Student and family variables	
INDIGENOUS	1= parent or guardian reports that indigenous language (or languages) are used to communicate in the home; 0=not
QUECHUA	1=parent or guardian reports that Quechua is used to communicate in the home; 0=not
AYMARA	1=parent or guardian reports that Aymara is used to communicate in the home; 0=not
FEMALE	1=female; 0=male
RURAL	1=located in rural area; 0=located in urban area
EDMOTHER	Series of dummy variables indicating educational attainment of mother, ranging from 0 (did not finish primary education) to 5 (finished higher education). EDMOTHER0 is excluded in regressions.
EDFATHER	Series of dummy variables indicating educational attainment of father, ranging from 0 (did not finish primary education) to 5 (finished higher education). EDFATHER0 is excluded in regressions.
BREAKFAST	1=student always has breakfast; 0=not
SEWERHOME	1=student's family has sewer connection; 0=not
ELECTHOME	1=student's family has electricity; 0=not
PHONE	1=student's family has telephone; 0=not
School and teacher variables	
INDTEACH	1=teacher's mother tongue is indigenous; 0=not
FEMTEACH	1=teacher is female; 0=male
AGETEACH	Age of teacher
EDTCH1	1=teacher has teaching degree; 0=not
EDTCH2	1=teacher is tenured without degree; 0=not
EDTCH3	1=teacher is "intern"; 0=not
BLKBRD1	1=classroom always has blackboard; 0=not
CHALK	1=classroom always has chalk; 0=not
LIBRARY	1=availability and condition of classroom library is average or better; 0=not
TEXT	1=more than half students have textbooks; 0=not
GUIDE	1=teacher guides are available; 0=not
PRIVATE	1=school is privately managed (but not by "convenio"); 0=not
CONVENIO	1=school is privately managed, but receives public funds; 0=not
BATHSCH	1=condition of school bathrooms or latrines is average or better; 0=not
ELECTSCH	1=condition of school electricity connection is average or better; 0=not
SEWERSCH	1=condition of school sewer connection is average or better; 0=not
COMPSCH	1=condition of school computer room is average or better; 0=not

Source: SIMECAL 1997.

BEST COPY AVAILABLE

6. The Distribution of Family and School Characteristics

This section explores how family and school characteristics are distributed between indigenous and non-indigenous students, between boys and girls, and between rural and urban students. It does so first by presenting basic descriptive statistics. It then presents results from probit regressions that explore the correlates of each family or school variable.

6.1 Family Characteristics

Table 4 reports descriptive statistics for the family and student variables that are used in the analysis. Overall, 30 percent of third-grade students and 22 percent of sixth-grade students live in families where an indigenous language is used to communicate. About two-thirds of these are Quechua speaking, and the rest Aymara (with, as noted, a small number of Quechua-Aymara bilinguals).

Table 4. Descriptive Statistics, Student and Family Variables

	Third-grade Sample					*Sixth-grade Sample*				
	Overall	*Spanish only*	*Indigenous*	*Quechua*	*Aymara*	*Overall*	*Spanish only*	*Indigenous*	*Quechua*	*Aymara*
SPANISH	49.65	50.59	47.47	47.85	46.89	51.66	52.75	47.73	47.98	47.69
	(9.31)	(9.25)	(9.07)	(9.40)	(8.43)	(10.61)	(10.60)	(9.71)	(9.52)	(10.11)
MATH	53.03	53.79	51.25	51.62	50.54	47.86	48.63	45.06	45.31	44.72
	(9.22)	(9.09)	(9.29)	(9.40)	(9.04)	(10.16)	(10.26)	(9.30)	(9.33)	(9.18)
INDIGENOUS	0.30	0.00	1.00	1.00	1.00	0.22	0.00	1.00	1.00	1.00
QUECHUA	0.19	0.00	0.64	1.00	0.13	0.15	0.00	0.69	1.00	0.16
AYMARA	0.12	0.00	0.41	0.08	1.00	0.08	0.00	0.37	0.09	1.00
FEMALE	0.52	0.53	0.48	0.49	0.47	0.49	0.50	0.48	0.51	0.43
RURAL	0.39	0.28	0.63	0.68	0.52	0.25	0.19	0.47	0.44	0.50
EDMOTHER0	0.13	0.09	0.23	0.28	0.14	0.12	0.09	0.23	0.25	0.18
EDMOTHER1	0.41	0.35	0.55	0.50	0.64	0.33	0.30	0.46	0.42	0.53
EDMOTHER2	0.16	0.18	0.13	0.12	0.14	0.21	0.22	0.19	0.18	0.21
EDMOTHER3	0.16	0.20	0.06	0.06	0.06	0.17	0.20	0.07	0.08	0.06
EDMOTHER4	0.10	0.13	0.03	0.03	0.02	0.12	0.14	0.04	0.05	0.02
EDMOTHER5	0.04	0.05	0.01	0.01	0.01	0.04	0.05	0.01	0.01	0.01
EDFATHER0	0.06	0.05	0.11	0.14	0.05	0.05	0.04	0.09	0.11	0.05
EDFATHER1	0.32	0.27	0.44	0.47	0.39	0.26	0.22	0.40	0.41	0.35
EDFATHER2	0.21	0.20	0.23	0.21	0.28	0.24	0.23	0.28	0.26	0.32
EDFATHER3	0.21	0.25	0.14	0.11	0.18	0.22	0.24	0.15	0.13	0.20
EDFATHER4	0.12	0.14	0.06	0.05	0.08	0.13	0.15	0.07	0.07	0.06
EDFATHER5	0.08	0.10	0.02	0.03	0.01	0.09	0.11	0.02	0.03	0.02
BREAKFAST	0.79	0.81	0.73	0.71	0.78	0.80	0.79	0.82	0.81	0.84
SEWER	0.29	0.35	0.17	0.19	0.15	0.38	0.41	0.26	0.30	0.20
ELECTRICITY	0.62	0.70	0.44	0.43	0.48	0.73	0.78	0.57	0.57	0.58
PHONE	0.17	0.22	0.06	0.07	0.05	0.22	0.26	0.07	0.08	0.06
Observations	10,534	7,352	3,182	2,037	1,310	11,215	8,789	2,426	1,672	902

Source: SIMECAL 1997.

BEST COPY AVAILABLE

The other descriptive statistics suggest three broad patterns, for both third and sixth graders. First, indigenous students are considerably more likely to live in rural areas; 28 percent of Spanish monolinguals live in rural areas, compared with 63 percent of indigenous students. Second, the parental education of indigenous students is markedly lower than that of Spanish monolinguals. Third, indigenous students live in households that are much less likely to receive basic services such as sewer, electricity, and telephone.

Table 5 assesses whether these patterns still hold in multivariate analyses. The table reports the results from 16 probit regressions, where each family variable is specified as a dependent variable. The results for the mother's education (EDMOTHER0 to EDMOTHER5) suggest that both indigenous students and rural students live in families where mothers have less schooling. For example, an indigenous student is 10 percent more likely to have a mother with that did not complete primary schooling, while a rural student is 11 percent more likely (see the results for EDMOTHER0). There is no gender gap, and there appear be few interactions between indigenous status, gender, and rural/urban location. A generally similar pattern holds for father's education.

The other striking pattern to emerge is the stark difference in availability of basic services to indigenous, but particularly rural students. For example, a rural student is 42 percent less likely to have access to electricity in the family home. An indigenous student, all else equal, is 8 percent less likely. A student who is both rural and indigenous is an additional 7 percent less likely. Again, there are no gaps between male and female students.

The pattern of results is suggestive that indigenous students live in families of lower socioeconomic status. This much is consistent with other empirical evidence from Bolivia. However, it suggests that this is not explained entirely by their greater propensity to live in rural areas. Even holding location constant, indigenous students live in families where the parents have less formal education, and the households are less privileged.

Table 5. Distribution of Student and Family Variables Among Groups, Probit Regressions

	EDMOTHER0	*EDMOTHER1*	*EDMOTHER2*	*EDMOTHER3*	*EDMOTHER4*	*EDMOTHER5*
INDIGENOUS	0.102**	0.207**	-0.029*	-0.104**	-0.081**	-0.030**
	(0.013)	(0.019)	(0.013)	(0.010)	(0.011)	(0.006)
FEMALE	-0.012	-0.018	-0.004	0.010	0.011	0.003
	(0.006)	(0.012)	*0.008)	(0.006)	(0.007)	(0.004)
INDIGENOUS*FEMALE	0.003	0.027	0.001	0.001	-0.009	0.001
	(0.010)	(0.016)	(0.014)	(0.014)	(0.012)	(0.007)
RURAL	0.107**	0.173**	0.006	-0.100**	-0.081**	-0.044**
	(0.014)	(0.020)	(0.012)	(0.010)	(0.012)	(0.008)
INDIGENOU*RURAL	-0.013	-0.139**	-0.037*	-0.017	-0.005	0.036*
	(0.014)	(0.019)	(0.016)	(0.017)	(0.017)	(0.017)
	EDFATHER0	*EDFATHER1*	*EDFATHER2*	*EDFATHER3*	*EDFATHER4*	*EDFATHER5*
INDIGENOUS	0.033**	0.166*	0.046**	-0.067**	-0.057**	-0.064**
	(0.009)	(0.019)	(0.015)	(0.013)	(0.010)	(0.012)
FEMALE	-0.004	-0.009	-0.022*	0.009	0.008	0.010
	(0.004)	(0.010)	(0.009)	(0.007)	(0.006)	(0.007)
INDIGENOUS*FEMALE	0.003	0.017	0.007	0.005	-0.015	-0.004
	(0.007)	(0.015)	(0.014)	(0.014)	(0.012)	(0.011)
RURAL	0.068**	0.223**	0.023	-0.106**	-0.068**	-0.091**
	(0.008)	(0.019)	(0.013)	(0.012)	(0.012)	(0.014)
INDIGENOUS*RURAL	-0.008	-0.105**	-0.036*	-0.016	0.008	0.036
	(0.008)	(0.018)	(0.016)	(0.021)	(0.017)	(0.021)
	BREAKFAST	*SEWER*	*ELECTRICITY*	*PHONE*		
INDIGENOUS	0.036**	-0.077**	-0.082**	-0.133**		
	(0.014)	(0.027)	(0.022)	(0.019)		
FEMALE	-0.006	0.030	0.011	0.016		
	(0.011)	(0.018)	(0.011)	(0.012)		
INDIGENOUS*FEMALE	-0.017	0.008	0.008	0.022		
	(0.015)	(0.019)	(0.017)	(0.017)		
RURAL	0.025	-0.342**	-0.417**	-0.224**		
	(0.016)	(0.031)	(0.026)	(0.025)		
INDIGENOUS*RURAL)	-0.128**	0.013	-0.074*	0.034		
	(0.027)	(0.038)	(0.037)	(0.033)		

Note: Each variable in columns 2-7 was regressed on the five independent variables in column 1. Marginal effects from each probit regression are reported (evaluated as a discrete change in the independent variables from 0 to 1). Huber-White standard errors, adjusted for clustering within schools, are in parentheses. The number of observations of all regressions is 21, 749.

Source: SIMECAL 1997 and authors' calculations.

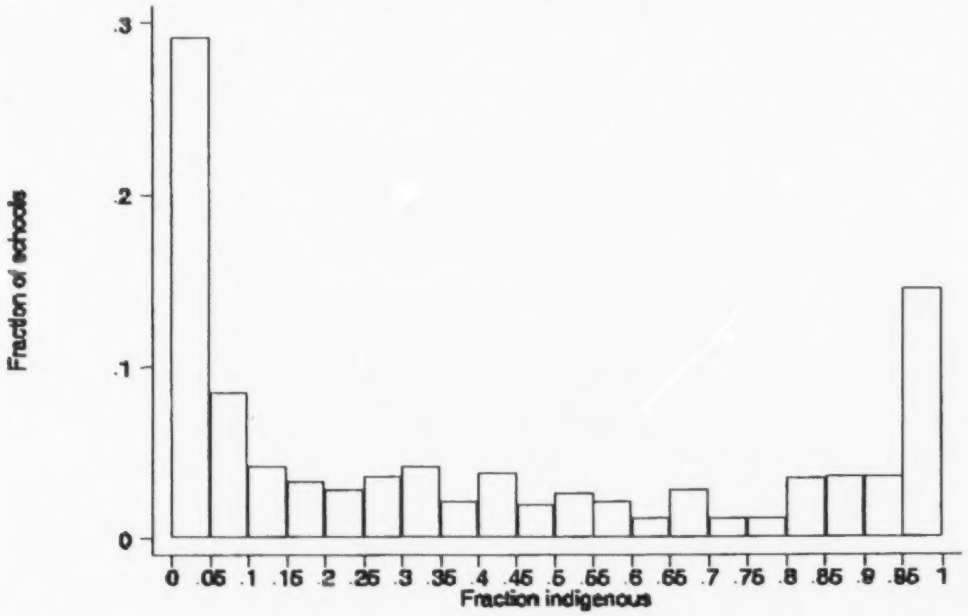
25 BEST COPY AVAILABLE

6.2 Segregation by Indigenous Status

Student achievement is a function of individual attributes, but it is also potentially a function of the attributes of other students that conform to school peer groups. In the present context, individual achievement may be affected by the linguistic background of other students in the same classroom or school.[19] Therefore, this section briefly examines how indigenous students are distributed across schools. It finds, perhaps not surprisingly, that they are distributed unevenly.

As Figure 1 demonstrates, just under 30 percent of the schools in the sample have between 0 and 5 percent indigenous students (that is, students who speak an indigenous language in the home). At the opposite extreme, 15 percent of schools are predominantly—over 95 percent—indigenous. The bimodal distribution is suggestive of rather intense segregation by linguistic status.

Figure 1: Distribution of Indigenous Students Across Primary Schools



Source: SIMECAL 1997 and authors' calculations.

As a further means of exploring this, we estimated a common measure of segregation, the exposure rate (E) of non-indigenous to indigenous students.[20] It is defined as $E = (1/N)\sum_k N_k [I_k / (N_k + I_k)]$, where N_k and I_k are the number of non-indigenous and indigenous

[19] The exact mechanism of this relationship is always straightforward to assess. For example, it may be the result of a student's exposure, or lack thereof, to Spanish when interacting with peers. It might also reflect the influence of teacher characteristics, if teachers alter their expectations or instructional methods in classrooms that are predominantly indigenous. For a review of the U.S. literature on peer effects, see McEwan (2000); for evidence of peer effects in Chile, see McEwan (2001).

[20] For a recent application to racial segregation in the U.S., see Clotfelter (2001).

BEST COPY AVAILABLE

students, respectively, in school k. N is the total for the entire sample. E provides an estimate of the proportion of indigenous students in the average non-indigenous student's school. In the present sample, that figure is 0.16, compared with a samplewide proportion of 0.27..

A further estimate of segregation is given by $S = (P - E)/P$, where P is the samplewide proportion of indigenous students. If $S = 0$, it implies perfect balance across schools in the exposure of indigenous and non-indigenous students, while $S = 1$ implies perfect segregation. In this sample, S is equal to 0.38.

Thus, segregation by linguistic status is notable in Bolivian primary schools. A full explanation is beyond the scope of this paper, but segregation is partly rooted in two factors. First, there is extensive residential segregation, with indigenous students more likely to live in rural areas and in marginal neighborhoods of urban areas. Patterns of residential sorting affect the choice of public schools. Second, a substantial proportion of students attend private schools. Indigenous students have lower incomes and, as might expected, are substantially less likely to attend these schools (this is explored further below).

Although later empirical analyses will not explicitly control for observed peer variables—such as the percentage of indigenous students in a school—they will do so indirectly via the inclusion of school and classroom fixed effects.

6.3 School and Teacher Characteristics

Table 6 reports the descriptive statistics for school and teacher variables. Overall, 23 percent of third graders and 16 percent of sixth graders have a teacher who reports an indigenous language as a "mother tongue." Indigenous students are more likely to have such a teacher (i.e., 38 percent of indigenous third graders, versus 16 percent of Spanish monolinguals). Notwithstanding this difference, it is still striking that more than 60 percent of indigenous students in the sample do not attend class with a teacher who is a mother-tongue speaker of an indigenous languages. Even when indigenous students have indigenous teachers, there is mismatch between language backgrounds. In this sample, roughly 5 percent of Quechua students are with a teacher whose mother tongue is Aymara. Similarly, about 10 percent of Aymara students are taught by mother-tongue speakers of Quechua. At the very least, this is cause for concern, presuming that teacher language ability is a necessary but not sufficient condition for effective bilingual instruction.

Table 6 also suggests fairly consistent differences between indigenous and non-indigenous students in access to basic instructional resources and school infrastructure. On average, indigenous students are less likely to attend classrooms with adequate libraries, textbooks, and teacher guides. They are substantially less likely to attend private schools (3 percent of indigenous third graders, versus 20 percent of non-indigenous third graders). The schools they do attend have worse bathroom facilities, electricity and sewer service, and computer facilities.

The multivariate analyses of Table 6 generally support these patterns, and they also suggest inequities related to urban/rural location. On average, indigenous students are 15 percent more likely to

have an indigenous teacher. With regards to other teacher characteristics, it appears that rural students are at a disadvantage. They are taught by younger teachers, on average, who are 16 percent less likely to possess a teaching degree. The distribution of teacher characteristics does not appear to be a function of whether students are indigenous.

With the important exception of textbooks, few of the differences in instructional materials are statistically significant. Both indigenous and rural students are less likely to attend classrooms where more than half the students have textbooks. These inequities are somewhat attenuated for student who are both indigenous and rural (indicated by the positive coefficient on the interaction term), although they are still at a net disadvantage relative to urban, non-indigenous students.

Rural students are considerably less likely to attend schools with adequate bathroom facilities, electricity and sewer service, and computer facilities (between 26 and 44 percent). To a lesser extent, this is also true for indigenous students, regardless of where they live. There are rarely inequities in the distribution of teacher or school resources across genders. Likewise, there are relatively few interactions between indigenous status, rural location, or gender.

Table 6. Descriptive Statistics, Student and Family Variables

	Third-grade Sample					*Sixth-grade Sample*				
	Overall	*Spanish only*	*Indigenous*	*Quechua*	*Aymara*	*Overall*	*Spanish only*	*Indigenous*	*Quechua*	*Aymara*
INDTEACH	0.23	0.16	0.38	0.30	0.51	0.16	0.12	0.30	0.28	0.32
FEMTEACH	0.72	0.78	0.61	0.69	0.50	0.59	0.61	0.53	0.56	0.49
AGETEACH	41.07	42.20	38.76	38.84	38.65	41.88	42.33	40.31	40.06	41.69
	(9.10)	(9.09)	(8.69)	(8.71)	(8.51)	(6.81)	(6.51)	(7.57)	(7.70)	(7.16)
EDTEACH1	0.85	0.85	0.85	0.88	0.80	0.94	0.95	0.88	0.88	0.90
EDTEACH2	0.07	0.07	0.07	0.05	0.11	0.01	0.01	0.03	0.04	0.00
EDTEACH3	0.08	0.08	0.08	0.07	0.09	0.05	0.04	0.10	0.09	0.10
BLKBRD	0.95	0.95	0.93	0.93	0.93	0.99	0.99	1.00	1.00	1.00
CHALK	0.85	0.87	0.80	0.82	0.74	0.86	0.87	0.84	0.91	0.63
LIBRARY	0.45	0.47	0.40	0.45	0.33	0.24	0.25	0.22	0.18	0.34
TEXT	0.49	0.53	0.41	0.41	0.42	0.49	0.53	0.33	0.41	0.19
GUIDE	0.91	0.93	0.87	0.83	0.92	0.97	0.96	0.98	1.00	0.95
CONVENIO	0.07	0.08	0.04	0.06	0.00	0.04	0.04	0.03	0.04	0.02
BATHSCH	0.74	0.78	0.64	0.66Q	0.62	0.74	0.79	0.57	0.57	0.59
ELECTSCH	0.67	0.76	0.49	0.51	0.47	0.79	0.84	0.65	0.67	0.62
SEWERSCH	0.45	0.54	0.25	0.26	0.24	0.59	0.63	0.46	0.48	0.40
COMPSCH	0.25	0.32	0.12	0.13	0.10	0.36	0.42	0.17	0.16	0.19
Observations	4,127	2,777	1,350	889	536	1,837	1,430	407	302	127

Note: Standard deviations are in parentheses for the continuous variable.

Source: SIMECAL 1997 and authors' calculations.

BEST COPY AVAILABLE

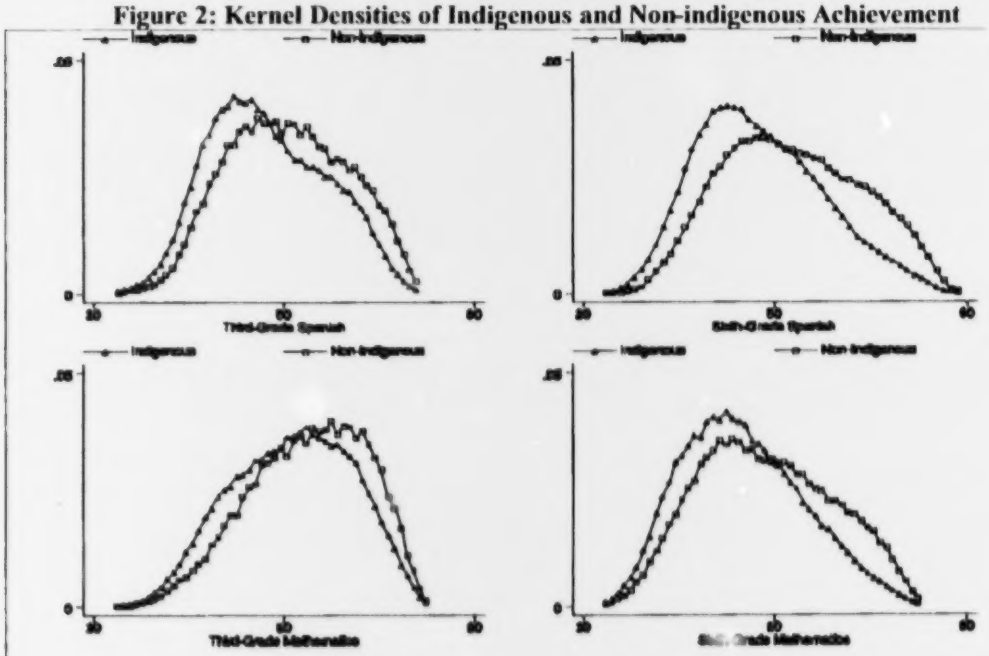
7. Patterns and Determinants of Achievement Differences

This section describes patterns in the distribution of academic achievement in Spanish and mathematics. It further assesses whether these differences in achievement can be explained by inequities in the distribution of family and school variables.

7.1 The Distribution of Academic Achievement

Figure 2 reports kernel densities that compare the distributions of indigenous and non-indigenous achievement in each subject and grade level. They suggest that the non-indigenous students, on average, obtain higher test scores in both grades. Nonetheless, there is a great deal of overlap in the distributions. While subsequent analyses focus on mean differences in achievement between each group, it should be emphasized that there is still considerable variance within each group.

Figure 2: Kernel Densities of Indigenous and Non-indigenous Achievement



Source: SIMECAL 1997 and authors' calculations.

Table 8 provides regressions results for the third-grade sample, using SPANISH as the dependent variable. For the moment, we shall restrict our attention to Models (1)-(3). Model (1) suggests that indigenous achievement is 3.1 points lower, on average, than non-indigenous achievement. With the addition of two additional variables (FEMALE and RURAL), the coefficient on INDIGENOUS declines by more than one third. Moreover, the coefficient on RURAL is strongly negative (-3.12), suggesting that part of the indigenous disadvantage is associated with location. The

BEST COPY AVAILABLE

coefficient on FEMALE is positive, but its magnitude is small compared to the others. Finally, there are no apparent interactions between these variables. That is, the gender gap is no larger or smaller among indigenous children; similarly, the indigenous gap is no larger or smaller in rural areas, compared with urban areas.

Table 7. Distribution of School and Teacher Variables Across Linguistic Groups, Probit Regressions

	INDTEACH	*FEMTEACH*	*AGETEACH*	*EDTEACH*	*EDTEACH2*	*EDTEACH3*
INDIGENOUS	0.145**	0.014	0.182	-0.037	0.038	-0.002
	(0.041)	(0.068)	(0.988)	(0.044)	(0.023)	(0.039)
FEMALE	0.026	0.093**	0.907	0.015	-0.007	-0.007
	((0.031)	(0.028)	(0.636)	(0.014)	(0.008)	(0.011)
INDIGENOUS* FEMALE	-0.054	-0.004	-1.184	0.012	-0.006	-0.005
	(0.028)	(0.032)	(0.657)	(0.018)	(0.012)	(0.014)
RURAL	0.064	-0.178*	-4.781**	-0.155*	0.076	0.078
	(0.064)	(0.088)	(1.486)	(0.065)	(0.052)	(0.046)
INDIGENOUS* RURAL	0.121	-0.121	-1.528	0.068	-0.043*	-0.012
	(0.068)	(0.099)	(1.546)	(0.039)	(0.019)	(0.042)
	BLKBRD	*CHALK*	*LIBRARY*	*TEXT*	*GUIDE*	*PRIVATE*
INDIGENOUS	-0.049*	-0.061	-0.020	-0.173**	-0.010	-0.196**
	(0.024)	(0.034)	(0.049)	(0.059)	(0.024)	(0.063)
FEMALE	0.005	0.014	0.036	0.014	0.030*	0.025
	(0.005)	(0.015)	(0.043)	(0.035)	(0.014)	(0.026)
INDIGENOUS* FEMALE	0.012	-0.021	-0.071	0.058	-0.010	0.006
	(0.007)	(0.021)	(0.037)	(0.038)	(0.019)	(0.035)
RURAL	0.011	-0.007	-0.039	-0.384**	-0.072	
	(0.025)	0.054)	(0.081)	(0.073)	(0.038)	—
INDIGENOUS* RURAL	0.031	0.013	0.066	0.282**	0.000	
	(0.019)	(0.043)	(0.084)	(0.076)	(0.032)	—
	CONVENIO	*BATHSCH*	*ELECTSCH*	*SEWERSCH*	*COMPSCH*	
INDIGENOUS	-0.007	-0.103	-0.177**	-0.193**	-0.094	
	(0.020)	(0.060)	(0.067)	(0.070)	(0.072)	
FEMALE	-0.005	0.005	-0.002	0.033	0.055	
	(0.016)	(0.024)	(0.026)	(0.037)	(0.035)	
INDIGENOUS* FEMALE	0.024	0.019	0.058*	0.097**	0.008	
	(0.015)	(0.026)	(0.026)	(0.035)	(0.043)	
RURAL	-0.037	-0.319**	-0.262**	-0.435**	-0.305**	
	(0.041)	(0.086)	(0.088)	(0.090)	(0.075)	
INDIGENOUS* RURAL	-0.053*	0.081	-0.032	0.026	-0.129*	
	(0.026)	(0.068)	(0.087)	(0.100)	(0.057)	

Source: SIMECAL 1997 and authors' calculations

Note: Each variable in columns 2-7 was regressed on the five independent variables in column 1. Marginal effects from each probit regression are reported (evaluated as a discrete change in the independent variables form 0 to 1). Huber-White standard errors, adjusted for clustering within schools, are in parentheses. The number of observations of all regressions is 5,964.

BEST COPY AVAILABLE

Table 8. Determinants of Third-Grade Spanish Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
INDIGENOUS	-3.12**	-1.93**	-2.71**	-1.08*	-0.97**	-0.68*
	(0.54)	(0.49)	(0.61)	(0.47)	(0.30)	(0.29)
FEMALE	—	0.57*	0.64*	0.35	0.26	0.18
		(0.25)	(0.31)	(0.25)	(0.17)	(0.16)
INDIGENOUS*FEMALE	—	—	-0.20	-0.07	0.20	0.12
			(0.43)	(0.40)	(0.29)	(0.28)
RURAL	—	-3.35**	-3.91**	-1.25	—	—
		(0.69)	(0.81)	(0.77)		
INDIGENOUS*RURAL	—	—	1.69	0.70	0.26	0.17
			(0.95)	(0.86)	(0.40)	(0.39)
EDMOTHER1	—	—	—	-0.58	-0.02	0.07
				(0.38)	(0.23)	(0.23)
EDMOTHER2	—	—	—	-0.37	0.13	0.15
				(0.44)	(0.29)	(0.28)
EDMOTHER3	—	—	—	0.91	0.65*	0.60
				(0.49)	(0.32)	(0.31)
EDMOTHER4	—	—	—	2.60**	1.67**	1.56**
				(0.56)	(0.37)	(0.37)
EDMOTHER5	—	—	—	3.93**	2.25**	1.99**
				(0.63)	(0.50)	(0.49)
EDFATHER1	—	—	—	-0.71	0.20	0.24
				(0.45)	(0.32)	(0.30)
EDFATHER2	—	—	—	-1.03*	0.39	0.41
				(0.49)	(0.35)	(0.33)
EDFATHER3	—	—	—	-0.44	0.60	0.55
				(0.49)	(0.37)	(0.35)
EDFATHER4	—	—	—	0.21	1.05**	1.09**
				(0.58)	(0.40)	(0.39)
EDFATHER5	—	—	—	1.54*	1.34**	1.37**
				(0.64)	(0.46)	(0.45)
BREAKFAST	—	—	—	0.85**	0.97**	0.70**
				(0.30)	(0.17)	(0.17)
SEWER	—	—	—	1.15**	0.07	0.10
				(0.42)	(0.21)	(0.20)
ELECTRICITY	—	—	—	1.14**	0.90**	0.89**
				(0.36)	(0.19)	(0.18)
PHONE	—	—	—	3.29**	0.48	0.27
				(0.45)	(0.26)	(0.25)
Constant	50.59**	51.24**	51.36**	48.02**	47.43**	47.64**
	(0.45)	(0.55)	(0.57)	(0.82)	(0.36)	(0.35)
R^2	0.02	0.05	0.05	0.14	0.52	0.57
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 10,534.

Tables 9 to 11 provide results for sixth-grade Spanish, third-grade mathematics, and sixth-grade mathematics. The results are generally similar, although two contrasts should be noted. First, the

BEST COPY AVAILABLE

negative coefficients on INDIGENOUS in the third grade become slightly larger in the sixth grade, although not by an order of magnitude. Second, girls have lower mathematics achievement than boys, although the results are only statistically significant in the sixth grade.

To better interpret the magnitude of these differences between indigenous and non-indigenous students, it is helpful to express them as effect sizes, or percentages, of a standard deviation.[21] This is done in Table 12 where each regression coefficient from Model (1) is divided by the standard deviation of its respective dependent variable (see the column of "unadjusted" differences). The effect sizes range between 0.28 and 0.47 of a standard deviation, with somewhat larger gaps in the sixth grade.[22]

To judge whether these are large or small, it is helpful to draw some comparisons. Consider that blacks in the United States currently score about one standard deviation lower than whites, on average (Hedges and Nowell 1998). Another useful comparison is found in Urquiola (2000), who uses the same data as this paper. He finds that an eight or nine student reduction in class size raises student achievement by an average of 0.17-0.45 standard deviations. Thus, the test score gaps between indigenous and non-indigenous students in Bolivia, while modestly large, are not overwhelming. At the very least, empirical evidence suggests that relatively simple interventions have potential to diminish the gap.

[21] In a standard "bell" curve, one standard deviation above and below the mean contains about 68 percent of the observations; two standard deviations above and below the mean contains about 95 percent.

[22] These are somewhat smaller than the gaps reported in Vera (1998), but that study's effects were driven by Aymara speakers in urban areas of La Paz and El Alto.

BEST COPY AVAILABLE

Table 9. Determinants of Sixth-Grade Spanish Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
INDIGENOUS	-5.01**	-3.56**	-3.21**	-0.97*	-0.40	-0.31
	(0.63)	(0.58)	(0.72)	(0.49)	(0.35)	(0.34)
FEMALE	—	0.51	0.74	0.47	0.40*	0.31
		(0.43)	(0.50)	(0.37)	(0.19)	(0.18)
INDIGENOUS*FEMALE	—	—	-1.05	-0.89	-0.85*	-0.69
			(0.58)	(0.50)	(0.37)	(0.36)
RURAL	—	-5.13**	-5.30**	-1.16	—	—
		(0.85)	(1.07)	(0.96)		
INDIGENOUS*RURAL	—	—	0.43	-0.88	0.43	0.32
			(1.29)	(1.15)	(0.51)	(0.50)
EDMOTHER1	—	—	—	-0.43	-0.88**	-0.80**
				(0.35)	(0.28)	(0.27)
EDMOTHER2	—	—	—	-0.39	-0.91**	-0.91**
				(0.42)	(0.32)	(0.31)
EDMOTHER3	—	—	—	0.69	-0.65	-0.58
				(0.49)	(0.36)	(0.35)
EDMOTHER4	—	—	—	3.36**	0.84*	0.96*
				(0.52)	(0.41)	(0.40)
EDMOTHER5	—	—	—	4.54**	0.90	1.08*
				(0.66)	(0.54)	(0.53)
EDFATHER1	—	—	—	0.51	0.97*	0.92*
				(0.50)	(0.39)	(0.38)
EDFATHER2	—	—	—	0.26	0.67	0.67
				(0.56)	(0.41)	(0.40)
EDFATHER3	—	—	—	1.29*	1.41**	1.29**
				(0.61)	(0.43)	(0.42)
EDFATHER4	—	—	—	2.10**	1.91**	1.91**
				(0.65)	(0.46)	(0.45)
EDFATHER5	—	—	—	4.35**	2.38**	2.18**
				(0.69)	(0.52)	(0.51)
BREAKFAST	—	—	—	0.05	0.33	0.20
				(0.32)	(0.20)	(0.20)
SEWER	—	—	—	1.63**	0.40	0.36
				(0.43)	(0.22)	(0.21)
ELECTRICITY	—	—	—	2.83**	2.45**	2.31**
				(0.38)	(0.22)	(0.21)
PHONE	—	—	—	4.03**	0.80**	0.89**
				(0.42)	(0.26)	(0.26)
Constant	52.75**	53.44**	53.36**	46.83**	48.47**	48.68**
	(0.54)	(0.61)	(0.64)	(0.75)	(0.43)	(0.42)
R^2	0.04	0.08	0.08	0.23	0.47	0.52
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 11,215.

BEST COPY AVAILABLE

Table 10. Determinants of Third-Grade Mathematics Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
INDIGENOUS	-2.54**	-1.46**	-2.24**	-0.81	-1.10**	-0.81**
	(0.49)	(0.46)	(0.62)	(0.50)	(0.30)	(0.30)
FEMALE	—	-0.20	-0.18	-0.44	-0.61**	-0.65**
		(0.25)	(0.30)	(0.24)	(0.17)	(0.17)
INDIGENOUS*FEMALE	—	—	-0.03	0.08	0.69*	0.58*
			(0.44)	(0.41)	(0.30)	(0.29)
RURAL	—	-3.15**	-3.66**	-1.34	—	—
		(0.66)	(0.75)	(0.71)		
INDIGENOUS*RURAL	—	—	1.55	0.62	0.05	0.06
			(0.91)	(0.82)	(0.43)	(0.41)
EDMOTHER1	—	—	—	-0.74	-0.24	-0.13
				(0.37)	(0.24)	(0.24)
EDMOTHER2	—	—	—	-0.46	-0.18	-0.14
				(0.44)	(0.30)	(0.30)
EDMOTHER3	—	—	—	0.58	0.28	0.33
				(0.46)	(0.33)	(0.32)
EDMOTHER4	—	—	—	2.32**	1.21**	1.16**
				(0.53)	(0.37)	(0.37)
EDMOTHER5	—	—	—	2.96**	1.32**	1.21*
				(0.64)	(0.49)	(0.48)
EDFATHER1	—	—	—	-1.04*	0.21	0.17
				(0.48)	(0.32)	(0.31)
EDFATHER2	—	—	—	-1.02	0.65	0.58
				(0.53)	(0.35)	(0.34)
EDFATHER3	—	—	—	-0.49	0.87*	0.81*
				(0.54)	(0.37)	(0.36)
EDFATHER4	—	—	—	-0.28	0.98*	0.93*
				(0.59)	(0.41)	(0.40)
EDFATHER5	—	—	—	0.90	1.14*	1.07*
				(0.62)	(0.47)	(0.46)
BREAKFAST	—	—	—	0.54	0.80**	0.63**
				(0.28)	(0.18)	(0.17)
SEWER	—	—	—	1.30**	0.32	0.31
				(0.40)	(0.22)	(0.20)
ELECTRICITY	—	—	—	0.78*	0.76**	0.72**
				(0.34)	(0.20)	(0.19)
PHONE	—	—	—	2.87**	0.36	0.17
				(0.39)	(0.25)	(0.24)
Constant	53.79**	54.80**	54.93**	52.66**	51.63**	51.77**
	(0.40)	(0.49)	(0.50)	(0.79)	(0.37)	(0.36)
R^2	0.02	0.04	0.04	0.11	0.47	0.53
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 10,534.

BEST COPY AVAILABLE

Table 11. Determinants of Sixth-Grade Mathematics Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
INDIGENOUS	-3.57**	-2.86**	-3.09**	-1.42**	-0.41	-0.36
	(0.59)	(0.55)	(0.64)	(0.51)	(0.34)	(0.34)
FEMALE	—	-0.86*	-0.94*	-1.14**	-0.86**	-0.93**
		(0.39)	(0.45)	(0.36)	(0.19)	(0.19)
INDIGENOUS*FEMALE	—	—	0.35	0.47	0.08	0.07
			(0.55)	(0.50)	(0.36)	(0.36)
RURAL	—	-2.57**	-2.61**	0.45	—	—
		(0.84)	(1.04)	(0.97)		
INDIGENOUS*RURAL	—	—	0.16	-0.89	0.07	0.12
			(1.26)	(1.18)	(0.51)	(0.51)
EDMOTHER1	—	—	—	-0.84*	-0.93**	-0.82**
				(0.39)	(0.28)	(0.27)
EDMOTHER2	—	—	—	-0.76	0.86*	-0.82*
				(0.45)	(0.31)	(0.31)
EDMOTHER3	—	—	—	0.07	-0.86*	-0.76*
				(0.48)	(0.35)	(0.35)
EDMOTHER4	—	—	—	2.14**	0.11	0.32
				(0.51)	(0.41)	(0.41)
EDMOTHER5	—	—	—	3.73**	0.58	0.84
				(0.69)	(0.55)	(0.55)
EDFATHER1	—	—	—	0.52	0.80*	0.65
				(0.54)	(0.39)	(0.38)
EDFATHER2	—	—	—	0.10	0.39	0.26
				(0.59)	(0.40)	(0.40)
EDFATHER3	—	—	—	0.74	1.05*	0.88*
				(0.61)	(0.43)	(0.42)
EDFATHER4	—	—	—	1.14	1.33**	1.28**
				(0.65)	(0.47)	(0.46)
EDFATHER5	—	—	—	3.21**	1.63**	1.43**
				(0.68)	(0.52)	(0.52)
BREAKFAST	—	—	—	0.08	0.40*	0.33
				(0.34)	(0.20)	(0.20)
SEWER	—	—	—	1.06**	0.26	0.19
				(0.39)	(0.21)	(0.21)
ELECTRICITY	—	—	—	2.09**	1.79**	1.72**
				(0.37)	(0.21)	(0.21)
PHONE	—	—	—	3.40**	0.40	0.41
				(0.47)	(0.26)	(0.26)
Constant	48.63**	49.54**	49.58**	45.15**	46.27**	46.47**
	(0.48)	(0.55)	(0.56)	(0.80)	(0.43)	(0.43)
R2	0.02	0.03	0.03	0.13	0.42	0.47
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 10,534.

BEST COPY AVAILABLE

Table 12. Summary of Achievement Gaps Between Indigenous and Non-digenous Students, Expressed in Standard Deviations

	Unadjusted	*Adjusted for:*		
		Student/family controls	*School fixed effects*	*Classroom fixed effects*
Third-grade Spanish	-0.34*	-0.12*	-0.10*	-0.07*
		(35%)	(31%)	(22%)
Sixth-grade Spanish	-0.47*	-0.09*	-0.04	-0.03
		(19%)	(8%)	(6%)
Third-grade mathematics	-0.28*	-0.09	-0.12*	-0.09*
		(32%)	(43%)	(32%)
Sixth-grade mathematics	-0.35*	-0.14*	-0.04	-0.04
		(40%)	(11%)	(10%)

Source: Table 8-11, Models (1), (4), (5), and (6).

Notes: * indicates that the difference is statistically significant at 5 percent. Regression coefficients are divided by the standard deviation of the dependent variable to obtain reported effect sizes. Percentages are obtained by dividing the adjusted effect sizes in each row by the corresponding unadjusted effect size.

7.2 The Determinants of Academic Achievement

This section is devoted to assessing the determinants of achievement gaps. Once again, let us focus our attention on the third-grade Spanish results in Table 8. Model (4) adds controls for family and student background. In general, these variables are of the expected sign. Parental education is positively associated with achievement, but mainly at the highest levels. Students that regularly eat breakfast and those with better access to household services have higher achievement. The causal interpretation of these variables is unclear, however, since they may proxy unobserved variables such as family income. Also likely is that they partly reflect school and teacher variables that are correlated with the socioeconomic status of families.

The inclusion of family and student variables leads to a decline in the indigenous gap of about two thirds from its magnitude in Model (1). The extremely large rural disadvantage is no longer statistically significant. Finally, the slight advantage of females is no longer statistically significant, but its magnitude is not substantially different.

The addition of school fixed effects further diminishes the indigenous gap, but only slightly. (Note that rural gaps can no longer be estimated, because this variable is absorbed by the fixed effects.) Many of the coefficients on family control variables decline substantially in magnitude, however, suggesting that they were proxying for school variables prior to the inclusion of fixed effects. The addition of classroom fixed effects again diminishes the coefficient on INDIGENOUS by a small amount, suggesting that classroom variables explain a small part of the test score gap. However, the coefficient is still negative and statistically significant.

These results are roughly similar for third-grade mathematics (although females have lower achievement than males in this case). The sixth-grade results suggest even more strongly that the indigenous gaps are largely explained by the unequal distribution of family and school variables. For both Spanish and mathematics, the indigenous gap is no longer statistically significant in Model (6).

BEST COPY AVAILABLE

Table 12 summarizes the results from each table, expressing the regression coefficients on indigenous as fractions of a standard deviation. The size of the indigenous achievement gap, conditional on a full set of family and school controls, is 0.04-0.09 standard deviations (but only statistically significant for third graders). This unexplained portion represents between 6 and 32 percent of the original gaps.

7.3 Comparing Quechua and Aymara Students

The prior analyses included Quechua and Aymara students in the same category. To examine whether this is justified, Tables 13 and 14 reestimate the regressions using two variables (QUECHUA and AYMARA) instead of the more general INDIGENOUS.

Estimates from Model (1) appear to suggest that the third-grade achievement of Aymara students is slightly worse than Quechua students. However, the differences between the coefficients on QUECHUA and AYMARA are not statistically significant at 5 percent (as gauged by a Wald test). This holds for all dependent variables. Tentatively, we conclude that Aymara and Quechua students are similarly disadvantaged, on average.

Table 13. Alternative Specifications of the Determinants of Spanish Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Third-grade Spanish						
QUECHUA	-2.40**	-1.08	-1.00	-0.22	-0.58	-0.38
	(0.66)	(0.59)	(0.68)	(0.57)	(0.39)	(0.38)
AYMARA	-3.33**	-2.66**	-3.59**	-1.53*	-0.95*	-0.65
	(0.61)	(0.59)	(0.83)	(0.73)	(0.40)	(0.39)
FEMALE	—	0.56*	0.63*	0.34	0.28	0.19
		(0.25)	(0.31)	(0.25)	(0.17)	(0.16)
AYMARA*FEMALE	—	—	0.32	-0.17	0.03	-0.05
			(0.50)	(0.47)	(0.33)	(0.32)
QUECHUA*FEMALE	—	—	-0.05	0.06	0.28	0.25
			(0.58)	(0.57)	(0.40)	(0.39)
RURAL	—	-3.50**	-3.87**	-1.20	—	—
		(0.68)	(0.80)	(0.76)		
QUECHUA*RURAL	—	—	0.35	0.15	-0.02	-0.00
			(1.08)	(1.00)	(0.48)	(0.46)
AYMARA*RURAL	—	—	1.98	0.48	0.77	0.40
			(1.18)	(1.11)	(0.57)	(0.54)
R^2	0.02	0.05	0.06	0.14	0.52	0.57
Student and family controls?	No	No	No	Yes	Yes	Yes
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes
Dependent variable: Sixth-grade Spanish						
QUECHUA	-4.29**	-3.13**	-2.76**	-1.06	-0.73	-0.55
	(0.63)	(0.60)	(0.78)	(0.58)	(0.41)	(0.41)
AYMARA	-4.25**	-2.80**	-2.26*	-0.05	0.56	0.48
	(0.94)	(0.90)	(0.88)	(0.68)	(0.51)	(0.50)
FEMALE	—	0.52	0.73	0.46	0.37	0.28
		(0.43)	(0.50)	(0.37)	(0.19)	(0.18)
AYMARA*FEMALE	—	—	-1.39*	-1.17*	-0.53	-0.49
			(0.61)	(0.53)	(0.43)	(0.42)
QUECHUA*FEMALE	—	—	-0.01	0.16	-0.91	-0.67
			(0.83)	(0.76)	(0.56)	(0.54)
RURAL	—	-5.26**	-5.34**	-1.17	—	—
		(0.85)	(1.06)	(0.95)		
QUECHUA*RURAL	—	—	0.75	-0.03	0.69	0.65
			(1.27)	(1.16)	(0.59)	(0.59)
AYMARA*RURAL	—	—	-0.97	-2.60	-0.33	-0.41
			(1.93)	(1.77)	(0.78)	(0.76)
R^2	0.03	0.08	0.08	0.23	0.47	0.52
Student and family controls?	No	No	No	Yes	Yes	Yes
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 10,534. The number of observations in sixth-grade regressions is 11,215.

BEST COPY AVAILABLE

Table 14. Alternative Specifications of the Determinants of Mathematics Achievement, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Third-grade mathematics						
QUECHUA	-1.90**	-0.69	-0.73	-0.07	-1.16**	-0.98**
	(0.59)	(0.54)	(0.68)	(0.61)	(0.38)	(0.37)
AYMARA	-2.99**	-2.41**	-3.24**	-1.41	-0.82*	-0.47
	(0.62)	(0.61)	(0.82)	(0.72)	(0.40)	(0.39)
FEMALE	—	-0.22	-0.21	-0.46	-0.61**	-0.65**
		(0.25)	(0.29)	(0.24)	(0.17)	(0.17)
AYMARA*FEMALE	—	—	0.12	0.25	0.69*	0.63
			(0.51)	(0.50)	(0.34)	(0.33)
QUECHUA*FEMALE	—	—	-0.22	-0.14	0.51	0.42
			(0.63)	(0.60)	(0.43)	(0.41)
RURAL	—	-3.27**	-3.59**	-1.28	—	—
		(0.65)	(0.75)	(0.71)		
QUECHUA*RURAL	—	—	0.19	-0.06	-0.09	0.05
			(1.01)	(0.96)	(0.49)	(0.48)
AYMARA*RURAL	—	—	1.92	0.58	0.82	0.70
			(1.21)	(1.15)	(0.65)	(0.63)
R2	0.02	0.04	0.04	0.11	0.47	0.53
Student and family controls?	No	No	No	Yes	Yes	Yes
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes
Dependent variable: Sixth-grade mathematics						
QUECHUA	-2.96**	-2.36**	-1.89**	-0.64	-0.33	-0.33
	(0.60)	(0.57)	(0.76)	(0.65)	(0.40)	(0.40)
AYMARA	-3.37**	-2.72**	-3.69*	-2.04**	-0.22	-0.04
	(0.87)	(0.88)	(0.70)	(0.59)	(0.50)	(0.50)
FEMALE	—	-0.87*	-0.92*	-1.12**	-0.85**	-0.92**
		(0.39)	(0.44)	(0.36)	(0.19)	(0.19)
AYMARA*FEMALE	—	—	-0.19	-0.04	0.16	0.13
			(0.65)	(0.60)	(0.42)	(0.41)
QUECHUA*FEMALE	—	—	0.79	0.94	-0.12	-0.20
			(0.68)	(0.65)	(0.54)	(0.53)
RURAL	—	-2.65**	-2.63**	0.45	—	—
		(0.83)	(1.03)	(0.96)		
QUECHUA*RURAL	—	—	-0.82	-1.50	0.11	0.22
			(1.24)	(1.18)	(0.58)	(0.58)
AYMARA*RURAL	—	—	1.22	-0.01	-0.06	-0.21
			(1.82)	(1.74)	(0.81)	(0.78)
R2	0.02	0.03	0.03	0.13	0.42	0.47
Student and family controls?	No	No	No	Yes	Yes	Yes
School fixed effects?	No	No	No	No	Yes	No
Classroom fixed effects?	No	No	No	No	No	Yes

Source: SIMECAL 1997 and authors' calculations.

Notes: ** (*) indicates statistical significance at 1 percent (5 percent). Huber-White standard errors are in parentheses; in models without fixed effects, standard errors are adjusted for clustering within schools. The number of observations in all regressions is 10,534. The number of observations in sixth-grade regressions is 11,215.

BEST COPY AVAILABLE

7.4 Interpretations

More than two-thirds of the modestly large gaps between indigenous and non-indigenous students are explained by the unequal distribution of family and school variables. A substantial rural/urban gap is entirely explained by family variables (although the latter may also proxy school variables). Finally, a disadvantage for girls in mathematics, though small in magnitude, could not be well explained by any of the variables.

In any empirical analysis such as ours, it is hardly straightforward to decompose the contribution of individual variables (particularly when school and teacher variables were controlled for with fixed effects). However, the descriptive analyses suggested likely candidates for the test score gap between indigenous and non-indigenous students.

First, indigenous families are considerably more likely to live in impoverished circumstances. Parents have less schooling and their households have diminished access to basic services, all proxies for family income and the home educational environment. These variables are correlated positively with academic achievement. Moreover, indigenous students are often concentrated in the same schools. If peer-group effects are important, this could further reinforce the achievement gap between indigenous and non-indigenous students. This was not directly tested, but such effects would have been subsumed by the school and classroom fixed effects.

Second, indigenous families attend schools with fewer instructional materials and worse infrastructure (even controlling for the rural or urban location). Their classrooms are endowed with fewer textbooks, for example, a basic input that has been linked to achievement in a range of empirical studies.[23] Similarly, the quality of their school buildings is worse, and they are substantially less likely to attend private schools.

Even controlling for a wide array of variables, there are small gaps remaining in achievement, particularly for third graders. There are three potential interpretations of the remaining gap. First, it is entirely possible that individual traits such as socioeconomic status are not fully observed and that these unobserved variables are distributed unequally across indigenous and non-indigenous students. This cannot be tested with existing data.

Second, it is possible that indigenous and non-indigenous students receive different resources, *even within the same classroom*. The classroom fixed effects control for all classroom resources that are constant within classrooms. However, teachers may treat different students differently; in particular, indigenous students may be expected to perform at lower levels and held to lower standards.[24]

Third, it is possible that indigenous and non-indigenous students receive exactly the same resources within classrooms, but that these resources are differentially effective. Precisely because indigenous students have unique linguistic backgrounds, instructional materials and pedagogies designed

[23] For a review of nonexperimental evidence, see Fuller and Clarke (1994); for an experimental study, see Jamison, Searle, Galda, and Heyneman (1981).

[24] Ferguson (1998) analyzes the issue of teacher expectations in the context of the black-white test score gap in the U.S.

41 BEST COPY AVAILABLE

for monolingual Spanish students may be less effective. Again, this cannot be directly tested with current data.

8. Summary and Conclusions

Among other findings, this paper has reported test score gaps between indigenous and non-indigenous students of 0.28-0.47 standard deviations. These gaps are largely, but not entirely explained by the unequal distribution of family and school variables across these students.

In this context, what can be done to ameliorate the test score gap? There appears to be considerable scope for applying relatively straightforward solutions. Every classroom, for example, should have a minimal package of instructional materials like textbooks. Similarly, Urquiola (2000) has shown that reductions in class size produce effect sizes that are similar in magnitude to existing test score gaps.

Beyond these simple approaches, there is probably great potential to improve indigenous achievement via the implementation of bilingual education programs. These are doubly attractive because they may pursue multiple goals at once. Besides improving achievement in Spanish and mathematics, they may improve literacy in the indigenous language. This presumes that bilingual programs are not simply designed as a "bridge" to the dominant language. While this was the case in early pilot programs in Bolivia, the current education reform in Bolivia emphasizes the maintenance of the indigenous language as a key goal.

All that said, it should be emphasized that these conclusions are necessarily preliminary. There is very little rigorous evaluation of whether bilingual education programs—or, indeed, any large-scale education reform—have succeeded in improving academic outcomes in Latin America. Even less effort has been devoted to analyzing the costs of individual programs or reforms, in order to determine the most cost-effective means of attaining a particular educational goal.

As a final point, one should recall that indigenous achievement has always been interpreted relative to that of nonindigenous students. This should not imply that nonindigenous achievement is satisfactory. In the early grades, Bolivia scores below the regional mean of 11 Latin American countries (Willms and Somers 2001). Bringing indigenous students to parity with others is merely the first leg of a longer journey.

References

Albó, X. 1988. "Bilingualism in Bolivia." In C.B. Paulston (Ed.), *International handbook of bilingualism and bilingual education* (pp. 85-108). New York: Greenwood Press.

———. 1995a. *Bolivia plurilingüe: Guia para planificadores y educadores (Vol. 1)*. La Paz: UNICEF-CIPCA.

———. 1995b. *Bolivia pluralingüe: Guia para planificadores y educadores (Vol. 2)*. La Paz: UNICEF-CIPCA.

Angrist, J.D., and A.B. Krueger. 1999. "Empirical strategies in labor economics." In 0. Ashenfelter and D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1277-1366). Amsterdam: Elsevier.

Briggs, L.T. 1979. "A critical survey of the literature on the Aymara language." *Latin American Research Review* 14(3): 87-105.

———. 1982. "Bilingual education in Bolivia." In A. Miracle (Ed.), *Bilingualism.: Social issues and policy implications*. Athens: University of Georgia Press.

———. 1985. "Bilingual education in Peru and Bolivia." In N. Wolfson and J. Manes (Eds.), *Language of inequality*. Berlin: Mouton.

Chiswick, B.R., H.A. Patrinos, and M.E. Hurst. 2000. "Indigenous language skills and the labor market in a developing economy: Bolivia." *Economic Development and Cultural Change* 349-367.

Clotfelter, C.T. 2001. "Are whites still fleeing? Racial patterns and enrollment shifts in urban public schools, 1987-1996." *Journal of Policy Analysis and Management* 20(2): 199-221.

Comita,s, L. 1972. "Education and social stratification in contemporary Bolivia." In T.J. LaBelle (Ed.), *Education and development: Latin America and the Caribbean* (pp. 363-378). Los Angeles: University of California Press.

Ferguson, R.F. 1998. "Teachers' perceptions and expectations and the black-white test score gap." In C. Jencks and M. Phillips (Eds.), *The black-white test score gap* (pp. 273-317). Washington, DC: Brookings Institution Press.

Fergusson, D.M., M Lloyd, and L.J. Horwood. 1991. "Family ethnicity, social background and scholastic achievement—an eleven year longitudinal study." *New Zealand Journal of Educational Studies* 26(1): 49-63.

Fuller, B. and P. Clarke. 1994. "Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy." *Review of Educational Research* 64(1): 119-157.

Grinevald, C. 1998. "Language endangerment in South America: A programmatic approach." In L.A. Grenoble and L.J. Whaley (Eds.), *Endangered languages: Current issues and future prospects* (pp. 124-159). Cambridge: Cambridge University Press.

Hedges, L.V. and A. Nowell. 1998. "Black-white test score convergence since 1965." In C. Jencks and M. Phillips (Eds.), *The black-white test score gap* (pp. 149-181). Washington, DC: Brookings Institution Press.

Hornberger, N. 2000. "Bilingual education policy and practice in the Andes: Ideological paradox and intercultural possibility." *Anthropology and Education Quarterly* 31(2): 173-201.

Hornberger, N.H. and K.A. King. 1996. "Language revitalization in the Andes: Can the schools reverse language shift?" *Journal of Multilingual and Multicultural Development* 17(6): 427-441.

Jamison, D., B. Searle, K. Galda, and S. Heyneman. 1981. "Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement." *Journal of Educational Psychology* 73(4): 556-567.

Jencks, C. and M. Phillips. 1998. *The black-white test score gap*. Washington, DC: Brookings Institution Press.

Jimenez, W. and P.J. McEwan. 2001. *Indigenous students and primary school access in Bolivia*. Unpublished manuscript, UDAPE and University of Illinois at Urbana-Champaign.

Kelley, J. 1988. "Class conflict or ethnic oppression? The cost of being Indian in rural Bolivia." *Rural Sociology* 53(4): 399-420.

Lavy, V. 1998. "Disparities between Arabs and Jews in school resources and student achievement in Israel." *Economic Development and Cultural Change* 175-192.

López, L.E. 1997. "To Guaranize: A verb actively conjugated by the Bolivian Guaranis." In N.H. Hornberger (Ed.), *Indigenous literacies in the Americas: Language planning from the bottom up* (pp. 321-353). Berlin: Mouton de Gruyter.

McEwan, P.J. 2000. "The potential impact of large-scale voucher programs." *Review of Educational Research* 70(2): 103-149.

———. 2001. *Peer effects on student achievement: Evidence from Chile*. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Mizala, A., P. Romaguera, and T. Reinaga. 1999. *Factores que inciden en el rendimiento escolar en Bolivia*. Serie económica no. 61, Centro de Economía Aplicada, Universidad de Chile, Santiago.

Muñoz, H. 1997. *De proyecto a política de estado: La educación intercultural bilingüe en Bolivia, 1993*. La Paz: Universidad Pedagógica Nacional, GTZ, UNICEF-Bolivia.

Patrinos, H.A. 2000. "The cost of discrimination in Latin America." *Studies in Comparative International Development* 35(2): 3-17.

Patrinos, H.A. and G. Psacharopoulos. 1996. "Socioeconomic and ethnic determinants of age-grade distortion." *International Journal of Educational Development* 16(1): 3-14.

Plaza, P. and X. Ibó. 1989. "Educación bilingüe y planificación lingüística en Bolivia." *International Journal of the Sociology of Language* 77: 69-91.

Psacharopoulos, G. 1993. "Ethnicity, education, and earnings in Bolivia and Guatemala." *Comparative Education Review* 37(1): 9-20.

Rivarola, D.M., G. Corvalán, and L. Zuniga. 1977. *Determinantes del rendimiento educativo en el Paraguay*. Asunción: Centro Paraguayo de Estudios Sociológicos.

Romaine, S. 2000. *Language in society: An introduction to sociolinguistics* (2nd ed.). Oxford: Oxford University Press.

SIMECAL. 2000. *Rendimientos escolares: Educación intercultural bilingüe, lenguaje L1, L2 y matemáticas, 3ro y 5to de primaria* (Tech. Rep.). La Paz: SIMECAL, Ministerio de Educación.

Urquiola, M. 2000. *Identifying class size effects in developing countries: Evidence from rural schools in Bolivia*. Unpublished manuscript, Cornell University.

Urquiola, M., W. Jiménez, M.L. Talavera, and W. Hernany. 2000. *Los maestros en Bolivia: Impacto, incentivos y desempeño*. La Paz: Universidad Católica Boliviana.

Vera, M. 1998. *Bilingüismo y rendimiento escolar en Bolivia*. Unpublished M.A. thesis, ILADES-Georgetown University.

Willms, J.D. and M.A. Somers. 2001. "Family, classroom and school effects on children's educational outcomes in Latin America." *School Effectiveness and School Improvement.*

Wölck, W. 1973. "Attitudes toward Spanish and Quechua in bilingual Peru." In R.W. Shuy and R.W. Fasold (Eds.), *Language attitudes: Current trends and prospects.* Washington, DC: Georgetown University Press.

Wood, B. and H.A. Patrinos. 1994. "Urban Bolivia." In G. Psacharopoulos and H.A. Patrinos (Eds.), *Indigenous people and poverty in Latin America: An empirical analysis* (pp. 55-96). Washington, DC: World Bank.

World Bank. 2001. *Peruvian education at a crossroads: Challenges and opportunities for the 21st century*. Washington, DC: World Bank.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").